



PARtecipazione e CONflitto

*** The Open Journal of Sociopolitical Studies**

<http://siba-ese.unisalento.it/index.php/paco>

ISSN: 1972-7623 (print version)

ISSN: 2035-6609 (electronic version)

PACO, Issue 11(2) 2018: 497-510

DOI: 10.1285/i20356609v11i2p497

Published in July 15, 2018

Work licensed under a Creative Commons Attribution-Non commercial-Share alike 3.0 Italian License

SYMPOSIUM/3

WHEN CONTEXT MATTERS

Analyzing conflicts with the use of big textual corpora from Russian and international social media

Svetlana S. Bodrunova

St. Petersburg State University

1. Introduction. Zombies as a side of the conflict

In February 2013, I was going quickly through a dataset collected from the Russian-language Twitter starting from a range of hashtags and keywords describing a violent conflict in the Moscow district of Biryuliovo. During this conflict, the local Muscovites almost demolished a warehouse where hundreds of re-settlers from the post-Soviet Central Asia dwelled and traded, after an alleged killing of a local guy by an Azerbaijani. This conflict became the most widely discussed by the Russian media and on social media platforms out of a range of similar conflicts rising in various parts of Russia wherein a violent inter-personal trigger made deeper inter-group tensions show up, required the involvement of local authorities and, in some cases, yielded to changes in policymaking.

While I was running through the dataset, I discovered a dozen tweets saying in Russian:

Urgent! Breaking! Zombies in #Biryuliovo!!! Attacking everyone! They're everywhere!

My team considered these tweets like spam or a crazy joke, irrelevant to the analysis and thus eliminated them from the sample. However, our choice seemed not to be the best one as, approximately one month later, we realized that there were three main actors in the conflict we were researching upon – namely, *vegetables*, *zombies*, and *cosmonauts*. *Vegetables* were the warehouse settlers: the word referred simultaneously to the vegetables that were the main object of storage at the warehouse and to vegetative state of mind. *Zombies* were the host communities who protested – they were rising up as zombies and coming from different directions to the warehouse, crushing windows and arguing with the police in an unstoppable manner. And the police men, in spherical helmets, were *cosmonauts*, the third party that represented the distant state existing somewhere out there in an outer space and never available on request.

The zombies made us rethink the role of contextual knowledge for the analysis of big textual corpora, as these corpora oftentimes do not open up to researchers' eyes in an expected way. Building on our research experience, in the remainder of this commentary I will claim for bigger attention to contextual knowledge in current studies of political conflicts, as contextual factors may turn online discussions into distorted mirrors of the respective societies and, hence, are relevant in relation to both the production and the interpretation of results.

2. It is the context that matters

After looking at the Biryuliovo case, we analyzed inter-ethnic conflicts on Twitter not only in Russia but also across the world – such as the Ferguson riots, the Cologne harassment and the Christmas Berlin bus crash, the Charlie Hebdo massacre and the Paris and Brussels attacks. Amongst other aspects, we looked at political polarization and political representation in these conflicts starting from the actual texts of tweets. Following the classic work by Tajfel and Turner (1979), we expected that the polarization of the online discussion would reflect a triple structure: the minority, the pro-minority majority, and the (largest, perhaps) anti-minority majority. However, the divisions we found were strikingly different from what we expected, and they could not be explained if not in relation to the wider context of political divisions and in light of the structures of the media systems, use of social networking platforms by audience groups, and the traditions of civil society.

In particular, in several inter-ethnic conflicts that we analyzed which took place in Russia, the discourse was occupied by only two groups – pro-establishment and anti-

establishment nationalists – and the only difference between them was that the first group openly supported the anti-migrant bashings and linked them to the “rising Russia”, while the second one blamed the authorities whom they perceived as the continuation of the elites of 1990s who “stole the country”. No group or community formed a pro-migrant counterweight in these discussions. Such representation was the result of several colliding factors, none of which could be explained leaning only on the discussion itself or considering exclusively elements such as the then very liberal policing towards migration from the post-Soviet space (which was one of the reasons for the rising inter-ethnic tensions in various parts of the country). Other factors, external to policing had, in our opinion, a bigger explanatory power.

First, the political spectrum in Russia is virtually non-existent in the conventional Western terms - that is, the left-right divisions do not work as differentiators for social groupings the way they work in most West European countries. In the absence of clear rules of the game in a post-Communist de-classed society (Kordonsky 2008), social macro-divisions form along post-Soviet vs. hypothetical “Western” values (albeit this generalization overlooks the existing complexities of the Western social divisions) as well as along earning and consumption lifestyles (Zubarevich 2011; 2013). This deep fragmentation shapes the political system, creating a cleavage between the “systemic” powers that are represented in the lower chamber of the Russian Parliament, the State Duma, and the non-systemic powers outside of it – a wide-ranging array of political actors, from the new Communists to liberal, monarchist, and anarchist movements, beyond the “liberal opposition” that is often recalled by media. Thus, no ‘left-right’ position clash could be expected to be found as sustaining the formation of majorities in a discussion over a socially polarizing issue, even a global one as immigration.

Second, there was no one that defended the migrants, as no NGOs or local authorities who could have balanced the anti-migrant voices were not active enough or, more precisely, completely absent from the discussion on Twitter. Similarly, no official representatives of the migrants were heard enough. This, in turn, was a clear reflection of the well-described weakness of the Russian civil society (Henderson 2011; Gelman and Ross 2010; Chebankova 2013). As to the liberally-oriented media whose discourse usually includes a human rights perspective, in these cases they mostly preferred to be neutral and informational or decided to pose only rhetorical questions to the authorities (Bodrunova *et al.* 2018).

Third, and perhaps most importantly, in all these conflicts the use of Twitter by the migrants in Russian was close to zero. First, they rarely communicated in Russian at all, and thus were missing from the sample. Second, in 2013, unlike in 2016 when the first scarce data on their use of mobile phones was obtained (Abezgil'din 2016), this com-

munity was, apparently, still using mostly button phones, not smartphones, and their use of computers was, as the same research suggests, lower in percentage compared to the Moscow host community. Thus, presumably, it was the structure of media consumption that, in fact, shaped the picture of the discussion on Twitter.

Here, we also need to note that platform use matters not only for the migrant community but *de facto* for any audience group in Russia. As many scholars have shown (Vartanova and Smirnov 2010; Bodrunova and Litvinenko 2015; Kiriya 2018), Russia has a peculiar social media market and audience structure. Thus, the main platform is not Facebook but *Vkontakte*, which catches a rather wide spectrum of audience's political views and values but is perceived by the scholars as more or less depoliticized and full of trashy information. In fact, a research conducted by the Laboratory of Internet Studies at Higher School of Economics has shown that the collections of *Vkontakte* texts are relatively poor in terms of substantial political talk (Kotlova and Koltcov 2013; Kotlova *et al.* 2017). The Russian-speaking segment of Facebook is a well-recognized echo chamber for 30+ liberal oppositional talk (Bodrunova and Litvinenko 2015; Kiriya 2018), even if there is lack of research that would bring on any percentages of the political views of its users. Similarly, the Russian-speaking *Livejournal* (livejournal.com) hosts tech-savvies and the writing and artistic community but, as we found out, a lot of conspiracy theories and biased 'patriotic' posts. *Odnoklassniki* ('Classmates', *odnoklassniki.ru* or *ok.ru*), instead, is a place for several audience groups, among them inhabitants of the rural areas of any age (but mostly older people and schoolchildren) and city inhabitants aged 50 or more. Even if there is a lack of data about the Russian-speaking segment of Twitter, there seems to be a different mode of representing politics than in the West. Politics is in fact discussed by media, nationalists/'patriots', and bots, and the latter may, in some cases, dominate (Stukal *et al.* 2017). That is, upon *none* of these platforms, the expected configuration of the conflicting groups could be found. Thus, the group structure of platform use is a factor that undermines hypotheses about social interaction that are based on theories developed after studying offline conflicts.

Fourth, the content of conventional media was never in favor of the migrant community. Several researches showed that print and online media, as well as the official discourse of the authorities, have significantly contributed to the formation of a one-sided view towards migrants as stupid, uneducated, and barbarian in general (Karpenko 2004; Mukomel' 2011; Malakhov 2015), while statistics tell us that many Central Asian incomers had higher education and the re-settler community was, on the overall, more educated than that they left behind (see Bodrunova *et al.* 2017). On federal TV channels, migrants were highly under-represented; even when they were discussed, the discourse on migration followed that of the authorities, controversial and rarely

consistent (Hutchins and Tolz 2015). Thus, neither media content available for retweeting could create any pro-migrant pole in the conflictual discussions.

Against this background, what could one expect to find in a Twitter dataset? Definitely, not what we had expected leaning on theory. Instead, we found a nationalist and radicalized anti-immigrant talk, “objective” voices of media bots and of “not-that-objective” political bots, and eyewitnesses of the bashings at the warehouse who tried to stay neutral but hardly could. Ultimately, in these cases of inter-ethnic conflicts, the Russian-speaking Twitter had become a pro-nationalist echo chamber that could not be balanced neither by the efforts of the official authorities (who, due to the clientelistic nature of Russian politics, are not interested in full and timely presence in microblogs) nor by those of the activist or of “ordinary users”.

Albeit substantively different, inter-ethnic conflicts we studied for Germany and the USA were clearly just as content-dependent as the Russian one. In the discussion on New Year’s Eve night’s mass harassment of women in the center of Cologne, we could see the classic left-right divisions. However, the anti-immigrant cluster was overwhelmingly bigger than the other and the pro-immigrant layer was led by left-leaning media like *Sueddeutsche Zeitung* or NGOs, thus reflecting not only the structure of the German civil society but also the level of institutional presence on Twitter. Moreover, we could find an intermediary layer of people who united the two clusters and made them talk to each other (Bodrunova *et al.* 2018).

The Ferguson case was, perhaps, the most surprising for us Russians, while it may not have been such for American media scholars. Here, the leaders of the discussion in the aftermath of the conflict were coming from the cultural sphere, like rap singers and priests. Moreover, the presence of the African-American community among the influencers was much higher than we would expect in accordance with population statistics, largely due to an intense use of Twitter by African Americans.

Ultimately, the interpretation of the political polarization and of the dominant discourses in all these inter-ethnic conflicts demanded a much wider knowledge on local political structures but also on media audiences and media content. Consistently, our expectations (including research questions and hypotheses) needed to be significantly corrected and context-bound, rather than coming directly out of the existing Western (or non-Western) theories. Eventually, our initial hypothesis never worked – the discussions were not about the minorities defending themselves and being defended against the views of the anti-minority part of the majority. Rather, online discussions were the distorted mirrors of the societies in which they were unfolding and were forming along deeper, more fundamental structural and values-based societal gaps. Perhaps, this conclusion may sound banal enough for social science researchers, but it

is often overlooked by data scientists with technical educational backgrounds. More relevantly, it was not only the social and political contexts that mattered. Also media, both in the form of content available for sharing and in the form of limitations of platform use by social groups, were very important in shaping the discussions. Without factoring in media and bringing in knowledge on them, our conclusions would have been very limited, if not completely misleading.

3. Theory-driven, data-driven, context-driven?

Ongoing scholarly discussion on data-driven social science fluctuates between the over-optimistic view on big-data as “new ‘hypothesis-neutral’ way of creating knowledge” (Mazzocchi 2015, 1250) and the defense of the theory- and hypotheses-driven research that has been there for centuries, as “proponents of ‘data-driven’ and ‘hypothesis-driven’ science argue over the best methods to turn massive amounts of data into knowledge” (Strasser 2012, 85).

In this debate, the opposition between theory-driven and data-driven social science seems to be overlooking the crucial importance of context as a separate third type of information that shapes research designs and helps re-interpret findings. Even the most influential conceptual papers on the challenges of big data analysis seem to overlook it (see, e.g., Fan *et al.* 2014; Burgess and Bruns 2012). This element demands more attention, as background information is often unavailable for the teams who try to do comparative studies and have good command of languages and history of the countries under scrutiny but not the current mediated discourses and peculiarities of political structures and cleavages.

Decades of research in social science and humanities have elaborated mechanisms of incorporation of background knowledge, like grounded theory formation or a wide array of other methodologies in qualitative studies. And there is no doubt that the criteria for rigorous qualitative knowledge should also apply to mixed-method research based on big datasets (Tracy 2010). In my opinion, it is the type of contextual knowledge that needs to be rethought, as, for political studies based on social media datasets, there are types of background information that next-to-always need to be taken into account.

Defining what context needs to be taken into account often mixes with other issues of data collection and analysis, and yet it poses specific questions. Accounting for context is not tantamount to the matter of replicability of existing studies, which many authors, including Bruns (2013), mention as a growing issue in communication studies.

Similarly, it is not adherent to the issue of the technical nature of the data available on social media platforms, which shapes and “sometimes limits the type of analyses that can be conducted” (Weller and Kinder-Kurlanda 2015: 29; see also Giglietto *et al.* 2012). It is different from the problem of personal biases and user features that distort and blur users’ views (Haggittai 2015) and it differs also from the problem of the limited access that academics have to data due to API affordances and other policies of the social media content providers (Puschmann and Burgess 2013) as well as from the distortions that come from how the data are extracted and datasets are made (Crawford 2013; Gitelman 2013). Conversely, accounting for context is closer to the issue of non-representativity (boyd and Crawford 2012) as well as to the problem inherent to the datasets from real-world social media, which have that “aura of truth, objectivity, and accuracy” wisely identified by boyd and Crawford (2012) and that creates in the researchers a feeling that the results will anyway be telling and show how the world really is. As Hitchin notes, data create the feeling of “all-seeing, infallible God’s eye view” (Hitchin 2014, 4; see also Haraway 1991). Tufekci (2014) draws attention towards the non-consideration of “field effects” – e.g. events external to online discussions (including even weather) that change users’ behaviors, for example, determining country to country variations within the Arab Spring. In other words, researchers often expect that data will provide all the explanations, and, with a proper research method, would reveal on their own the answers to research questions. But even if there are cases when we need to “let the data speak for themselves” (Gould 1981), the point is that, unlike in astronomy or genome decoding, the results in studies of social media do not just *live* in context. More often, what can be seen within data *is just* a product of context. Thus, the distortions in the data may be even bigger than boyd and Crawford (2012) have famously described, claiming that “taken out of context, data lose meaning and value” (p. 670).

In most cases, political studies of social big data (Manovich 2011; Bello-Organ *et al.* 2016) can – and should – be understood as context-driven. Not in the sense that context completely shapes the research hypotheses but in the sense that defining and exploring the relevant context needs to become a must to make sense of the datasets from social media, as well as to correct researchers’ expectations and interpretations.

However, there is another question raised by this need to take context into account: will the embedding of contextual knowledge put another layer of constraints upon our reasoning and hypothesizing, which is already narrow enough to make the conclusions in methodologically diverse papers non-comparable? Will contextualization further prevent generalization based on many isolated research results? To avoid this, the po-

litical studies community needs to elaborate what types of contexts and on which levels of analysis we need to work with.

I argue that the opposition between data-driven and hypotheses-driven research may be partly relieved by the use of context on different analytical levels similarly to what Pat Langley suggested for automated discovering of physical laws in massive data as early as in 1981 where “[t]he lowest levels correspond[ed] to direct observations, while the highest correspond[ed] to hypotheses that explain[ed] everything so far observed” (Langley 1981, 31). Contextual knowledge may be used to better link the existing epistemological levels, without bringing researchers to the point at which they need to reformulate epistemologies towards data-driven science, an approach against which many scholars have raised their concerns (Frické 2015).

4. Media-based contexts for cross-cultural research

Context is always there, but it is perceived as highly case-dependent. However, as I argued above, political studies of social big data need to take into account the broader context and, even, multiple contexts and contextual areas, insofar as contexts shape the content of social media as well as the links between users. In this sense, accounting for context works as a filter to tune too abstract hypotheses and too abstract data-driven expectations.

Context is a challenge – especially for cross-cultural projects, which often yields to reduce their analytic scopes and limit the range of research questions that can be answered. Nonetheless, this challenge can be transformed into a research opportunity, as exploration of relevant contexts may pose additional research questions that would allow for establishing the comparability of cases (for example, inter-ethnic conflicts) under scrutiny.

For instance, contextual areas would, perhaps, include cultural and linguistic aspects of national speech, or discursive taboos, or histories of the national discussions on a given topic. In the cases of the inter-ethnic conflicts that we were researching upon, political context such as the structure of civil society, political polarization and party spectra, was in fact very relevant.

As a media scholar, however, I would claim again to focus on media-related contexts, as media today themselves constitute a context of increasing importance but which has remained relatively unstructured so far.

Table 1 suggests some media-related aspects that our research group has found to be relevant in many projects in social media research on Russian, English, German, and

French-language datasets. In the table, I describe three domains of media-related context – i.e., the outer media system, the platforms and their audiences, and the case-specific context – for two stages of research – namely, preliminary exploration and interpretation of research findings. Essentially, this proposed approach argues that knowledge on media contexts allows not only for more precise formulation of research hypotheses and selection of proxies but also helps evaluate research results against the known contexts.

Table 1. Media as context: suggestions for political studies

Context domains	Research stages	
	<i>Pre-hypothesis exploration, formulation of hypotheses, and selection of proxies</i>	<i>Interpretation of results</i>
Outer media system	<ul style="list-style-type: none"> • Structure of the politically relevant media segments • General structure of media bias • Audience cleavages and groups in media consumption • Relative activity of media on social networking platforms • Online/offline media parallelism 	Evaluation of presence and relative importance of media discourse induced: <ul style="list-style-type: none"> - via content sharing - via presence of media among influencers
Platforms and platform audiences	<ul style="list-style-type: none"> • Competition structure on the market of social platforms • Profiles of average user • Amount of institutional presence (relevant media, political actors, commercial sector, NGOs etc.) • Amount of bots 	Evaluation of the results against: <ul style="list-style-type: none"> - the average user profile - amount of institutional presence - amount of bots
Case-specific contexts	<ul style="list-style-type: none"> • The role of particular media in formation of streamlines in a given discussion offline • Possible limitations introduced by the platforms (bans of users, algorithmic changes for a discussion etc.) 	Defining case differences and similarities linked to: <ul style="list-style-type: none"> - media roles in the discussions - platform providers' approaches

Source: Author's elaboration

Within a sort of an “onion” domain structure, the first domain describes the media system in a given country outside the platform under scrutiny, including how the media segments (newspapers, TV channels, online media etc.) are structured in terms of political bias, audience groups, online/offline presence, and activity in social media. This will

allow for assessing how exactly the existing media system can contribute to shape the discourse via two channels, namely content sharing by users and activity and popularity of media accounts.

The second domain focuses on the role of particular platforms and their typical audiences in forming a discussion, or, more precisely, on possible deviations of the audience profiles from population distributions. Here, one needs to consider existing data on the place of a given platform within the social media market, as well as the profile of its average user and/or major user groups. I also suggest that the actors whose activity significantly distorts the “natural” user talk are to be taken into account and, among them, particular attention should go to institutional and automated (bot-like) accounts.

Finally, the third domain deals with how media and/or social networking platforms act within a particular discussion. In many cases, media behave as a bee swarm leaning the public opinion to a particular direction, and some media may change the pace and direction of the discussion by publishing breaking news. Also, platforms introduce limitations on what can and cannot be discussed. Of course, for each particular project, evaluation of the suggested aspects is subject to available knowledge. This, in its turn, raises an issue of having enough research, both nationally and cross-culturally, on media and platforms, their content, markets, and audiences.

5. Conclusions

The analysis of how users talk about political conflicts led us to several conclusions on the necessity to embed more contextual (background) knowledge into studies based on social big data. To summarize the arguments stated above, I propose a list of key points to the argument of this commentary:

1. Societies still matter. Their political polarization, civil activism, media audiences, and discursive traditions shape online discussions to the extent that they become distorted mirrors of societies themselves.
2. This creates the necessity of taking context into account – and raises the question of what contexts need to be studied, and how exactly.
3. Contexts may be used as “reality filters” to tune both exclusively theory-driven and data-driven academic approaches. Moreover, they can also become a research space in their own right.
4. Within contextual factors, media-related aspects are of increasing importance. These include available media content to share, media bias within various segments of

media markets, the structure of the market of social-communicative platforms, and platforms, views, and values-based audience grouping;

5. Media scholars should be involved more often into political and social science research on political and social issues whenever social media data are used.

6. Zombies can be real.

Acknowledgements

This research was supported by Russian Science Foundation, research grant 16-18-10125.

References

- Bello-Orgaz G., J. J. Jung, D. Camacho (2016), "Social big data: Recent achievements and new challenges", *Information Fusion*, 28: 45-59.
- Bodrunova S. S., I. S. Blekanov, A. A. Litvinenko, and A.S. Smoliarova (2018), "Beyond left and right? Real-world political grouping in Twitter discussions on inter-ethnic conflicts in Germany, France, the USA, and Russia", *Media and Communication*, 2019, Vol.2 (special issue).
- Bodrunova S. S., O. Koltsova, S. Koltcov, and S. Nikolenko (2017), "Who's Bad? Attitudes Toward Resettlers From the Post-Soviet South Versus Other Nations in the Russian Blogosphere", *International Journal of Communication*, 11: 3242–3264.
- Bodrunova S. S., A. A. Litvinenko (2015), "Four Russias in communication: fragmentation of the Russian public sphere in the 2010s", in B. Dobek-Ostrowska, M. Glowacki (eds), *Democracy and Media in Central and Eastern Europe*, London: Peter Lang, pp. 63-79.
- boyd d., K. Crawford (2012), "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon", *Information, communication & Society*, 15(5): 662-679.
- Bruns A. (2013), "Faster than the speed of print: Reconciling 'Big Data' social media analysis and academic scholarship", *First Monday*, 18(10). Retrieved August 10, 2018 (<https://www.ojphi.org/ojs/index.php/fm/article/view/4879/3756>).
- Burgess J., A. Bruns (2012), "Twitter archives and the challenges of" Big Social Data" for media and communication research", *M/C Journal*, 15(5). Retrieved August 11, 2018 (<http://journal.media-culture.org.au/index.php/mcjournal/article/view/561Driscoll/0>).

- Chebankova E. (2013), *Civil society in Putin's Russia*, London: Routledge.
- Crawford K. (2013), "The hidden biases in big data", *Harvard Business Review*, April 1, 2013. Retrieved August 11, 2018 (<https://hbr.org/2013/04/the-hidden-biases-in-big-data>).
- Fan J., F. Han, and H. Liu (2014), "Challenges of big data analysis", *National Science Review*, 1(2): 293-314.
- Frické M. (2015), "Big data and its epistemology", *Journal of the Association for Information Science and Technology*, 66(4): 651-661.
- Giglietto F., L. Rossi, and D. Bennato (2012), "The open laboratory: Limits and possibilities of using Facebook, Twitter, and YouTube as a research data source", *Journal of Technology in Human Services*, 30(3-4): 145-159.
- Gitelman L. (ed.), *'Raw Data' is an Oxymoron*, Cambridge, MA: MIT Press.
- Gould P. (1981), "Letting the data speak for themselves", *Annals of the Association of American Geographers*, 71(2): 166-176.
- Haraway D. (1991), *Simians, Cyborgs and Women: The Reinvention of Nature*, New York: Routledge.
- Henderson S. L. (2011), "Civil society in Russia: state-society relations in the post-Yeltsin era", *Problems of Post-Communism*, 58(3): 11-27.
- Karpenko O. (2004), "What do migrants threaten and how? Discursive games towards 'guests from the South' and their consequences [Как и чему угрожают мигранты: Языковые игры в «гостей с юга» и их последствия]", in T. Baraulina & O. Karpenko (eds), *Migration and the nation-state* [Миграция и национальное государство], St. Petersburg: Center for Independent Social Research, pp. 62-84.
- Kiriya I. (2018), "Russian Media", in *Russia*, London: Palgrave Macmillan, pp. 97-103.
- Koltsova O., S. Koltcov (2013), "Mapping the public agenda with topic modeling: The case of the Russian Livejournal", *Policy & Internet*, 5(2): 207-227.
- Koltsova O., S. Nikolenko, S. Alexeeva, O. Nagornyy, and S. Koltcov (2017), "Detecting interethnic relations with the data from social media", in *Proceedings of International Conference on Digital Transformation and Global Society*, Cham: Springer, pp. 16-30.
- Kordonsky S. G. (2008), *The estate structure of the post-Soviet Russia* [Сословная структура постсоветской России], Moscow: Institut fonda 'Obschestvennoe mnenie', vol. 219.
- Langley P. (1981), "Data-driven discovery of physical laws", *Cognitive Science*, 5(1): 31-54.
- Malakhov V. S. (2015), *Integration of Migrants: Concepts and Practices* [Интеграция мигрантов: Концепции и практики], Moscow: Mysl'.

- Manovich L. (2011), "Trending: The promises and the challenges of big social data", *Debates in the Digital Humanities*, 2: 460-475.
- Mukomel' V. I. (2011), "Russian discourses on migration: the 'zero years'" [Российские дискурсы о миграции: "нулевые годы"], *Russia reforming itself* [Россия реформирующаяся], (10): 86-109.
- Mazzocchi F. (2015), "Could Big Data be the end of theory in science?: A few remarks on the epistemology of data-driven science", *EMBO reports*, 16(10): 1250-1255.
- Puschmann C., J. Burgess (2013), "The politics of Twitter data", HIG Discussion Paper Series, No. 2013-01. Retrieved August 12, 2018 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2206225).
- Strasser B. J. (2012), "Data-driven sciences: From wonder cabinets to electronic databases", *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(1): 85-87.
- Stukal D., S. Sanovich, R. Bonneau, and J. A. Tucker (2017), "Detecting bots on Russian political Twitter", *Big Data*, 5(4): 310-324.
- Tajfel H., J. C. Turner (1979), "An integrative theory of intergroup conflict", in W. G. Austin, S. Worchel (eds), *The social psychology of intergroup relations*, Monterey, CA: Brook/Cole Publishing (pp. 33-47).
- Tracy S. J. (2010), "Qualitative quality: Eight 'big-tent' criteria for excellent qualitative research", *Qualitative Inquiry*, 16(10): 837-851.
- Tufekci Z. (2014), "Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls", *ICWSM*, 14: 505-514.
- Vartanova E., S. Smirnov (2010), "Contemporary structure of the Russian media industry", in A. Rosenholm, K. Nordenstreng, E. Trubina (eds), *Russian mass media and changing values*, London: Routledge, pp. 21-40.
- Weller K., K. E. Kinder-Kurlanda (2015), "Uncovering the challenges in collection, sharing and documentation: The hidden data of social media research", in *Papers from the 2015 ICWSM Workshop on Standards and Practices in Large-Scale Social Media Research*, pp. 28-37).
- Zubarevich N. (2011), "Four Russias [Четыре России]", *Vedomosti*, December 30, 2011. Retrieved August 7, 2018 (www.vedomosti.ru/opinion/news/1467059/chetyre_rossii.html).
- Zubarevich N. (2013), "What the four Russias should expect [Чего ожидать четырем Россиям]", *Vedomosti*, 24 September. Retrieved August 7, 2018 (www.vedomosti.ru/opinion/news/16681621/chetyre-rossii-chtodalshe?full#cut).

Author's Information

Svetlana S. Bodrunova is habilitated Doctor of Political Science and Professor at School of Journalism and Mass Communications, St.Petersburg State University, Russia. She has published two books, several chapters, and over 50 research papers in Russian and English on Russian and European journalism, media and politics, social media, and ethnicity in communication. Her recent publications include papers published in *Journalism*, *Journalism Practice*, and *International Journal of Communication*, among others. Her research interests today lie in the area of studying social media via big data methodologies, hybridization of media systems, and Russian traditional and social media in comparative perspective. She leads Center for International Media Research in her university and collaborates as a guest researcher and expert with universities in Hamburg and Dortmund (Germany).